

Democratizando a Bioinformática

Kelly Hidalgo, Tiago Palladino, Victor Borin, Carla Paixão

2021-03-23

Contents

Prefacio	5
Por que ler/usar este livro?	7
Estrutura do livro	7
Informação sobre software e convenções	7

Prefacio





Esta é um guia **simples** e **didático** para se inciar no mundo da bioinformática para processamento e análises de sequenciamento em grande escala. No primeiro volume serão abordadas as análises de sequenciamento do amplicon do gene 16S rRNA pela plataforma *Illumina*.

Por que ler/usar este livro?

Este livro foi pensando para ajudar pessoas que tem pouca ou nenhuma experiência com bioinformática e no seus trabalhos ou projetos precisam analisar dados de microbiomas obtidos a partir de sequenciamento em larga escala, especialmente na plataforma Illumina. É um guia teórico/prático para processar este tipo de dados com total independência.

Foi criado a partir da experiencia dos autores, quem também tiveram que enfrentar dificuldades para iniciar com este tipo de análises e pensando em isto desenvolveram o presente guia para ajudar a novos iniciantes.

Estrutura do livro

O primeiro volume, está dividido em três grandes partes. A primeira tem uma introdução ao mundo das linguagens de programação, onde serão introduzidos os comandos mais básicos para se desenvolver na linha de comando. A segunda parte contempla o processamento de dados de sequenciamento do amplicon 16S rRNA, desde a obtenção dos dados, passando pelo controle de qualidade, a eliminação de quimeras, clusterização em ASV (*Amplicon sequence variant*) e anotação taxonômica, usando três ferramentas diferentes (*Qiime2*, *DADA2*, *Galaxy*) . Por último, são abordadas as análises para geração de gráficos e estatísticas utilizando principalmente a linguagem **R**, no entanto também são apresentadas outras alternativas.

Informação sobre software e convenções

O conteúdo de este livro será principalmente para usuários de Linux/MacOs. Para os usuários de Windows, se recomenda, instalar um programa ou aplicativo para aceder a um servidor com Linux (p.e. *MobaXterm*, *Putty*, *Secure Shell App de Google Chrome Apps*) ou instalar uma maquina virtual com Linux (recomendada **Ubuntu**). No entanto é sempre recomendável revisar a configuração do computador a ser usado e conferir que cumpre com os requisitos mínimos de cada software.

Os *pipelines* de este livro usam diversos software, que irão se apresentando na medida que sejam necessários. A maioria das ferramentas baseadas em código (*code based*), serão instaladas usando Bioconda's package de Anaconda.

O software **R** será bastante usado durante várias etapas do processamento e análises de dados. Se recomenda a instalação *R version 4.0.4* e o *IDLE RStudio* a versão mais recente. (**Nota:** A instalação destes software será abordada em seções mais para frente).

Convenções

Não será adicionados **prompts** (i.e. > e +) aos códigos em este livro. Os comentários dentro dos códigos serão sinalizados com **##**. Os nomes de ferramentas, softwares e pacotes serão escritos em negrita (p.e. **Qiime2**), e código na linha serão formatados em fonte typewriter (p.e. `qiime2 -h`).

Em construção...